Digital watermarking

J.-F. Delaigle, C. De Vleeschouwer, B. Macq

Laboratoire de Télécommunications et Télédétection
Université catholique de Louvain
Bâtiment Stévin - 2, place du Levant
B-1348 Louvain-la-Neuve
Tel.: +32 10 47.80.72 - Fax: +32 10 47.20.89
E-mail: delaigle@tele.ucl.ac.be

ABSTRACT

This paper presents a process able to mark digital pictures with an invisible and undetectable secrete information, called the watermark. This process can be the basis of a complete copyright protection system.
The process first step consists in producing a secrete image . The first part of the secret resides in a basic information that forms a binary image. That picture is then frequency modulated. The second part of the secret is precisely the frequencies of the carriers. Both secrets depends on the identity of the copyright owner and on the original picture contents. The obtained picture is called the stamp.
The second step consists in modulating the ampitude of the stamp according to a masking criterion stemming from a model of human perception. That too theoretical criterion is corrected by means of morphological tools helping to locate in the picture the places where the criterion is supposed not to match.
This is followed by the adaptation of the level of the stamp at that places. The so formed watermark is then added to the original to ensure its protection.
That watermarking method allows the detection of watermarked pictures in a stream of digital images, only with the knowledge of the picture owner's secrets.

Keywords: copyright protection, watermark, secrete key, masking, human vision model, perceptive components, morphology, robustness, detection, correlation.

1 GENERAL INTRODUCTION

With the increasing availability of digitally stored information and the development of new multimedia services, security questions are becoming even more urgent. The acceptance of new services depends on whether suitable techniques for the protection of the work providers' interests are available.[1]

Moreover the nature of digital media threatens its own viability:

- First the replication of digital works is very easy and, what is more dangerous, really perfect. The copy is identical to the original.

- The ease of transmission and multiple uses is very worrying, too. Once a single pirate copy has been made, it is instantaneously accessible to anyone who wants it, without any control of the original picture owner.
- Eventually the plasticity of digital media is a great menace. Any malevolent user *(a pirate)* can modify an image at will. Such maniplations are really easy for a pirate and put many copyright protection methods at risk.

According to these considerations the conception of a copyright protection system is really vital and it constitutes a great challenge, because it should cope with all these threats. Without watermarking, most authors will not dare to broadcast their work.

This paper presents an additive watermarking technique. It consits in producing a synthetic picture (also called the stamp) which holds informations about the ownership of the original image and depends on the picture contents. That stamp is added to the original in a way that resulting picture is perceptually identical to the original one and so that the stamp is indetectable by a pirate computer. The aim of that technique is not the authentication of the picture content nor the identication of the owner. It is to allow a controller (i.e. the owner's computer or a Trusted Third Part) to find out watermarked pictures in a stream of images with the knowledge of the owner's secret key in order to detect broadcast of illegal copies.

The most interesting part of that method is the embedding process i.e. the weighting of each pixels of the stamp before adding it to the original. This is based on the masking concept coming from a model of human vision (the perceptive model). From this concept was deduced a method which reveals itself actually efficient. Another interesting part is the presentation of two methods used for the detection of watermarked pictures without the original. This last point is fundemental for the management of the copyright protection.
Eventually this paper ends with the analyse of the results and the system robustness.

2 THE MASKING

2.1 Introduction

The aim of a watermarking technique is to provide an invisible embedding of a secrete information, the watermark. This watermark must be masked (hidden) by the picture it is inlayed in. Precisely a master thesis has lead to a masking criterion deduced from physiological and psychophysic studies.[2] Nevertheless, this theoretical criterion having been formulated for monochromatic signals, it had to be adaptated to suit real images.

2.2 The perceptive model: approximation of the eye functionment

It is now admitted that the retina of the eye splits an image in several components. These components circulate from the eye to the cortex by different tuned channels, one channel being tuned to one component.

The characteristics of one component are:

- the location in the visual field (in the image).
- the spatial frequency (in the Fourier domain: the amplitude in polar coordinates).
- the orientation (in the Fourier domain: the phase in polar coordinates)

So, one perceptive channel can only be excited by one component of a signal whose characteristics are tuned to its. Components that have different characteristics are independent.

2.3 The masking concept

According to perceptive model of human vision,[3] signals that have same (near) components take the same channels from the eye to the cortex. It appears that such signals interact and are submitted to non-linear effects. The masking is one of those effects.

Definition: *the detection threshold* is the minimum level below which a signal can not be seen.

Definition: *the masking* occurs when the detection threshold is increased because of the presence of another signal.

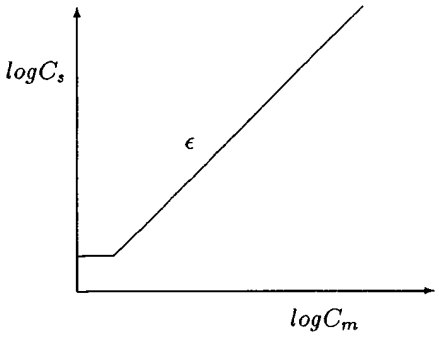
In other words, there is masking when a signal can not be seen because of another with near characteristics and at a higher level.

2.4 The masking model

With the object of modalizing the masking phenomenon, tests have been made on monochromatic signals, also called *gratings*. It appears that the eye is sensitive to the contrast of those gratings. This contrast is defined by:

$$C = \frac{2(Lmax - Lmin)}{Lmax + Lmin} \tag{1}$$

where L is the luminance.
It is possible to determine experimentally the detection threshold of one signal of contrast C_s with respect to the contrast C_m of the masking signal. That threshold can be modalized as follows:



Such bilogarithmic curves are traced for signals of one single frequency and one orientation (f_0, θ_0).
The expression of the detection threshold is thus:

$$C_s = max[C_0, C_0(\frac{C_m}{C_0})^{\epsilon}] \tag{2}$$

where ϵ (the slope) depends on (f_0, θ_0), typically, $0.6 \leq \epsilon \leq 1.1$.

It is possible to extend that expression to introduce frequency dependence. The general expression of the detection threshold is becomes:

$$C_s(C_m, f, \theta) = C_0 + k_{(f_0,\theta_0)}(f, \theta)[C_{s\,(f_0,\theta_0)}(C_m) - C_0] \tag{3}$$

where:

$$k_{(f_0,\theta_0)}(f, \theta) = exp[-(\frac{log^2(\frac{f}{f_0})}{F^2(f_0)} + \frac{(\theta - \theta_0)^2}{\Theta^2(f_0)}] \tag{4}$$

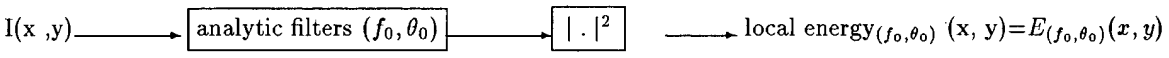
In that expression, f_0 and θ_0 are relevant to the masking signal, f and θ are relevant to the masked signal, $F(f_0)$ and $\Theta(f_0)$ are parameters that represent the spreading of the Gaussian function, C_0 is often negligable. The spread of the gaussian function depends upon the frequency f_0: For frequency, typical bandwith at half response are 2,5 octaves at 1 c/d and 1,5 octaves at 16 c/d with a linear decrease between both frequencies.[4] For orientation, half bandwith at half response depends on f_0 and it takes typical values like 30 degrees at 1 c/d and 15 degrees at 16 c/d.[5]

After this expression, the frequency dependence of the detection threshold has a Gaussian form. Only near frequency signals can interact. When the frequency of the masking signal (the mask) is far from this of the signal to mask, the detection threshold is almost equal to C_0.

2.5 The masking criterion

It is important to notice that those results concern only gratings signals. To deduce a masking criterion that will apply to signals like real images, the preceding masking condition has to be adaptated. So, it is necessary to define a new concept able to take the place of the contrast, because the contrast is not define for real images. That new concept,[2] is the *local energy.*

The local energy is defined on narrowband signals centered around one frequency and one orientation. A picture which is a broadband signal is first filtered by Gabor narrowband filters, whose characteristics are near to human perception. The local energy around one frequency and one orientation is calculated following the scheme presented in this figure:



The masking criterion: If the local energy of one picture is less than the local energy of the mask, around all the frequencies (f_0, θ_0) and for each pixel (x, y), then one can say that the picture is masked by the mask. Strictly, a picture is masked by a mask if $\forall(x, y)$ and $\forall(f_0, \theta_0), E_{mask,(f_0,\theta_0)}(x, y) \geq E_{picture,(f_0,\theta_0)}(x, y)$. For real images, a good approximation of this criterion can be obtained by using a bank of filters whose central frequencies correspond to independent components and which are spread on all the Fourier space. It is admitted that 4 or 5 frequencies and 4 to 9 orientations are sufficient. The standard choice is twenty filters (5 frequencies and 4 orientations).

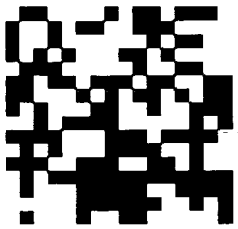


Figure 1: Example of basic information used

2.6 Conclusion

This section has lead to the expression of an easily implementable masking criterion appliable to any image. But this criterion is only an extension of a theoretic criterion appliable to monochromatic signals. Thus cases where that criterion does not match are possible.

3 PRINCIPLE OF THE SYSTEM

3.1 Basic information of the watermark

This information is a binary picture looking like a modified checkerboard (figure 1). As explained later, the pixels value of the square forming that picture can correspond to a binary sequence deduced from the copyright owner's (CO) *secrete key.*

3.2 The stamp

In order to take advantage of the eye behaviour, the basic information is modulated at different frequencies and orientations corresponding to rather independent components. Moreover, we take care to filter the initial checkerboard with a low pass filter (LPF) (i.e. a Butterworth LPF) so that the resulting signal is bandlimited. This point is very important because it permits to limit the verification of the masking criterion in the corresponding channel.
The position of the modulating carriers is *secret.* It can be deduced from CO's secret key. In practice, the frequency plan is divided into sectors. Each sector is relevant to one perceptive component and defined a group of couples (f, θ) where basic information can be modulated. Only one couple is chosen for each sector (because couples of a same sector don't stimulate independent components). The picture obtained from the sum of each modulated grid is called *the stamp* $S(x, y)$.

$$S(x,y) = \sum_{j \in K} G(x,y).cos(f_{x_j}.x + f_{y_j}.y) \quad (5)$$

K represents the set of sectors and (f_{x_j}, f_{y_j}) correspond to the couple chosen in sector j (this couple is designed by the CO's secrete key).

3.3 The position of the process in a global copyright scheme

The process should be placed in a copyright protection scheme like drawn at figure 2.
The skeletization function consists in an image processing program extracting essential characteristics from an image. The result is a bitsteam. This must be followed by a *hash-function*[6] whose result is a succession of blocks of bits. Every block has the same length. The skeletization function gives the same result for two near images (i.e. original image and watermarked image). But the H-function always gives different results from different bitstreams as inputs. So, the inscription keys will be different for perceptually distinct pictures. After the H-function, the ciphering function is a trapdoor function.[6] Thanks to this function the inscription keys used to deduce the basic grid and the position of the carriers depends on the CO's secret key.
The aim of the use of a trapdoor function is to prevent someone from reproducing the same inscription keys with the knowledge of the H-function result. But it is possible for anyone to inverse that trapdoor function and to find the H-function result from the inscription keys. It can be interesting in a proof procedure.

4 IMPLEMENTATION

4.1 Inscription

The purpose of the inscription is to adapt the level of each part of the stamp (for all frequencies) to make it invisible once added to the picture. As mentioned above, each part of the stamp is narrow band. Inscriptions at different frequencies are thus independent and one can treat the different components of the stamp one at a time. For each frequency designed by the inscription keys, the procedure is divided in three steps : the modulation, the regulation of the level and the correction.

- Modulation

  The first step consists in the modulation of the particular carrier by the lowpass grid $G(x,y)$. The result is $G(x,y).cos(f_{x_j}.x + f_{y_j}.y)$, where f_{x_j} and f_{y_j} are the carrier position.

- Regulation of the level

  According to the perceptual model, in order to guarantee the invisibility of the watermark its local energy has to be inferior to the picture local energy for each pixel around the inscription frequency. A way to reach this objective is to multiplicate the modulated grid by a weighting mask $Weight_j(x,y)$ reducing the amplitude of the stamp where energy in the correponding component of the original picture is weak. Nevertheless, one must take care to keep the narrow band characteristic of the resulting signal $S_j(x,y)$ (= $Weight_j(x,y).G(x,y).cos(f_{x_j}.x + f_{y_j}.y)$) in order to avoid non linear interactions between different parts of the stamp. In conclusion, $\forall j$, we have to find a signal $Weight_j(x,y)$ so that:

  - $\forall(x,y)\ E_{S_j}(x,y) < E_{I,(f_{xj},f_{xj})}(x,y)$
  - S_j is narrow band

  For simplification, lets consider $Weight_j(x,y)$ be composed of two factors:

  - α_j, a constant factor (fixing the global level of the stamp).
  - $M_j(x,y)$, a mask whose values $\in [0,1]$.

  When α_j is chosen, the way to find $M_j(x,y)$ so that $Weight_j(x,y)$ satisfy the conditions defined above is the following:

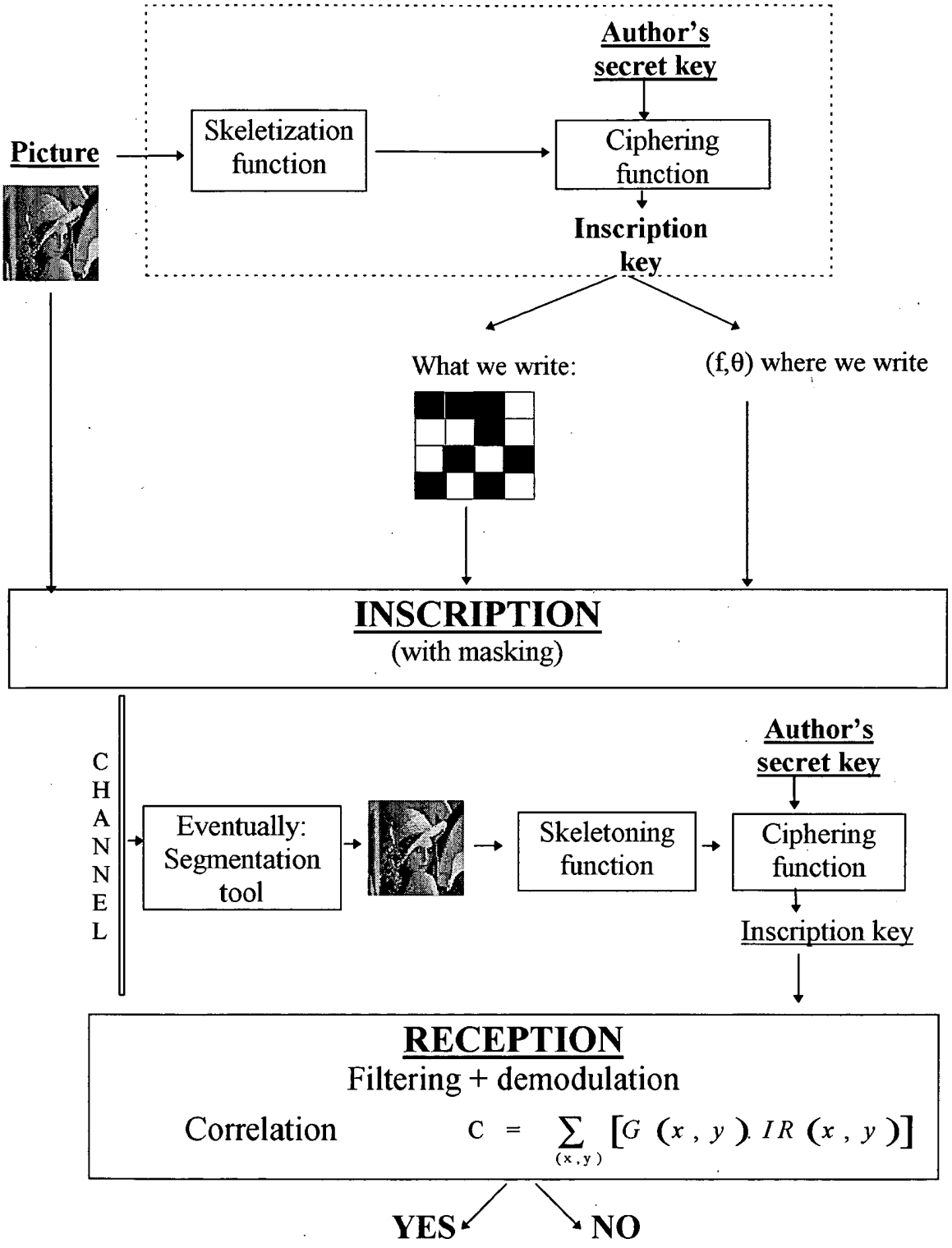
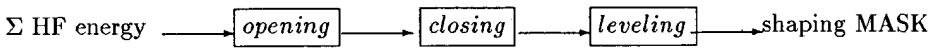


Figure 2: Global scheme for copyright protection.

- Firstly, $M_j(x,y)$ is a binary mask. $M_j(x,y) = 1$ when the local energy of the stamp pemits the masking and $M_j(x,y) = 0$ when the local energy of the stamp is too important. It is obvious that the initial choice of α_j has a direct influence on $M_j(x,y)$. Indeed, a great α_j value will lead to put most of the $M_j(x,y)$ values to zero, while a small α_j value will lead to keep most of $M_j(x,y)$ values at one.
- Secondly, $Weight_j(x,y)$ is filtered so that the stamp remains narrow band.
- After this second step, one has found a signal $\alpha_j.M_j(x,y).G(x,y)$ which is better masked than $\alpha_j.G(x,y)$. In order to really satisfy the masking criterion $\forall(x,y)$, this procedure must be repeated iteratively, taking $M_j(x,y).G(x,y)$ as new $G(x,y)$. Experiments have shown that only two iterations are sufficient to have a result satisfying the masking criterion everywhere.

One important question remains: how to choose α_j?
It has already been said that the more α_j increases, the more $M_j(x,y)$ has points equal to zero. A trade off has to be found by means of a defined criterion. Maximizing the correlation at the detection (by maximizing $\sum \alpha_j.M_j(x,y).G(x,y)$) could have been a good criterion, but such a criterion often tends to impose an optimum with a lot of points equal to zero and a small number of points with a great value. The addition of the so obtained watermark generally entails a degradation of the picture quality. This emphasizes the lack of the masking criterion used.
As mentioned in section 2.6, the invisibility criterion used here is an extension for real images. It appears that this extension entails some imperfections. This criterion being insufficient, some improvements have been brought thanks to experimental results.

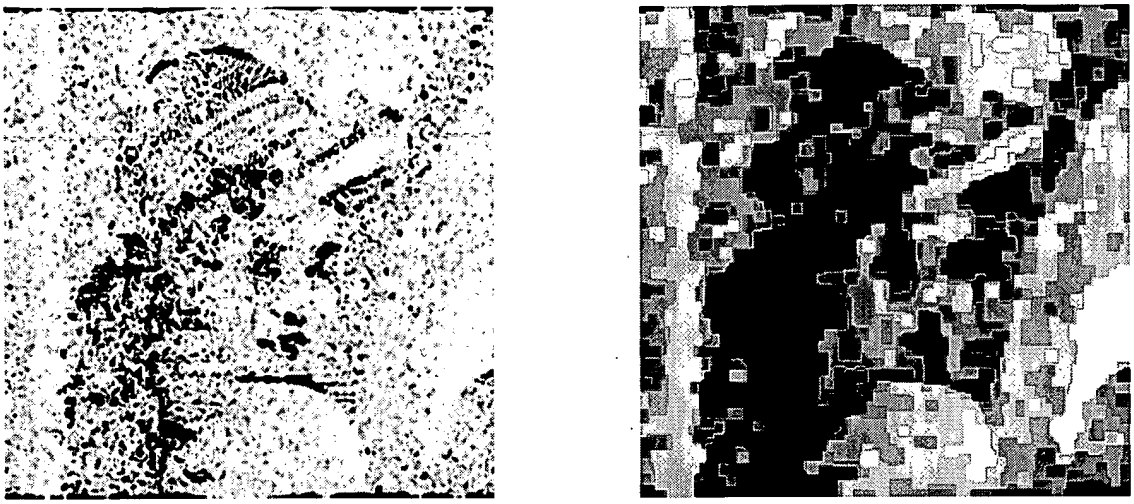
The conclusion of these observations is that the invisibility is only strictly observed in high activity regions, where the local energy of high fequencies is important. These regions have to be favoured during the inscription in the sense that the level of the watermark will be increased in those regions while it has to be decreased in other regions.
The correction process first isolates the high activity regions (figure 3.a). Then, an homogeneization of this picture is performed by use of morphological tools, e.g. one opening and one closing (figure 3.b). After a leveling (in fact, a division by the mean or mean square value of the homogenized mask), we obtain a new mask used to multiply the picture local energy and so, giving an advantage to regions of highfrequency energy in comparison with other areas. After that correction, the process is identical to the one described previously. Moreover, the complexity is not increased. Indeed, we first work on the inscription at high frequencies (where there is no quality problems). The value of high frequency local energy is then used for the calculation of the correcting mask used for inscription at lower frequencies. The correction scheme is drawn in the following schema.



4.2 Detection

The aim is to detect if a watermark has been embedded. This can be done with the use of a correlation, but first it is necessary to isolate the watermark and then to demodulate it in order to reconstruct something that is highly correlated with the basic information (the grid).
The formulation of the watermark is:

$$W(x,y) = \sum_{j \in K} A_j.cos(f_{x_j}.x + f_{y_j}.y) \tag{6}$$



(a) (b)

Figure 3: Correcting mask for Lena: (a) Areas of high frequencies, (b) Morphological homogeneization of the mask.

$$where A_j = \alpha_j . G(x,y) . M(x,y) \tag{7}$$

In this expression, $M(x,y)$ adjusts the level of the grid in order it becomes invisible, it is called *a mask*, and its maximal value is one.
α_j is a constant that used to normalize the mask, it must be as high as possible.

The detection is divided in three steps : teh demodulation, the correlation and the decision.

- Demodulation

$$I_W(x,y) = \sum_{j \in K} A_j . cos(f_{x_j} . x + f_{y_j} . y) + I_O + N(x,y) \tag{8}$$

where $I_W(x,y)$ is the watermarked picture, $I_O(x,y)$ is the original picture and $N(x,y)$ is an additive noise from the channel.

The demodulation consists in multiplying I_W by $cos(f_{x_j} . x + f_{y_j} . y), \forall j \in K$ and then to filter with a LP filter.
The result will be :

$$D_j(x,y) = \frac{1}{2} . A_j(x,y) + N^{\star}(x,y) \tag{9}$$

$N^{\star}(x,y)$ depends on the image and on the additive noise. The other parts of the stamp will be eliminated by the LP filter.

- Correlation It consists in mutiplying the demodulated information $D(x,y) = \sum_{j \in K} D_j(x,y)$ with the basic grid $G(x,y)$. If the picture has not been too deteriorated, $D(x,y)$ and $G(x,y)$ should be similar.

$$C = \sum_{j \in K} \sum_{x,y} D_j(x,y) . G(x,y) \tag{10}$$

$$= \sum_{j \in K} \alpha_j \sum_{x,y} [G^2(x,y).M_j(x,y) + G(x,y).N^\star(x,y)] \quad (11)$$

In 11, the first term is even greater than the second, because G and $N^\star$ have null average values.
So C exclusively depends on the watermark value.
in the case the grid is not the good one, the correlation gives:

$$C^\star = \sum_{j \in K} \alpha_j \sum_{x,y} G(x,y).G^\star(x,y)..M_j(x,y) \quad (12)$$

$C^\star \ll C$ if the choice of the basic information has been appropriate.

- decision

  The detection algorithm performs demodulations and correlations at diverse frequencies and with diverse grids.The decision is made after the comparison of these correlations.

5 RESULTS

The first and probably mosty important result is the invisibility of the stam in all images that were tested. Figure 4.a and b compares the original and stamped picture for Lena. In figure 4.e, omne observes the watermark that was added to the original picture.

Two methods were used to determine whether an image is watermarked or not. The first one consists in comparing the result of C the correlation made with the right grid $G(x,y)$ from the right key with $C^\star$ the correlation made with $G^\star(x,y)$, the grid obtained by random keys see 12. If the picture is watermarked, the correlation with the right key is even greater than the random correlations. The results below (Figure 5) show the pertinence of this method.

The second method uses a grid $G(x,y)$ formed from a MLS sequence, having good correlation properties. Correlations are made with shifted versions of the basic grid. Due to these good correlation properties, the correlation with the the right grid gives a result even greater than the correlations with shifted grids. Results are presented below (figure 4.c and d), if a picture is watermarked, a pick appears in the center.

6 SYSTEM ROBUSTNESS

Many tests have been performed concerning usual pictures deteriorations in image processing like blurring and compression. The inspection of these results are quite satisfying, but expected due to the frequency approach.
For all classical pirate attacks like zoom, cropping, overwatermarking it is not as simple. The overwatermarking makes no problem, the presence of the watermark is still detected. But for zoom and cropping, the remaing point is to find a few tools permitting to complete the process. The concept of these tools is already defined but yet no implementation has been acheived.[7]

7 CONCLUSION

The process developed here allows the watermarking of the ownership of any picture. The perceptual approach used here is probably the best one, that is why the results obtained are so satifying compared with other methods and this method is so performant. Nevertheless studies are still running to acheive a new goal, consisting in

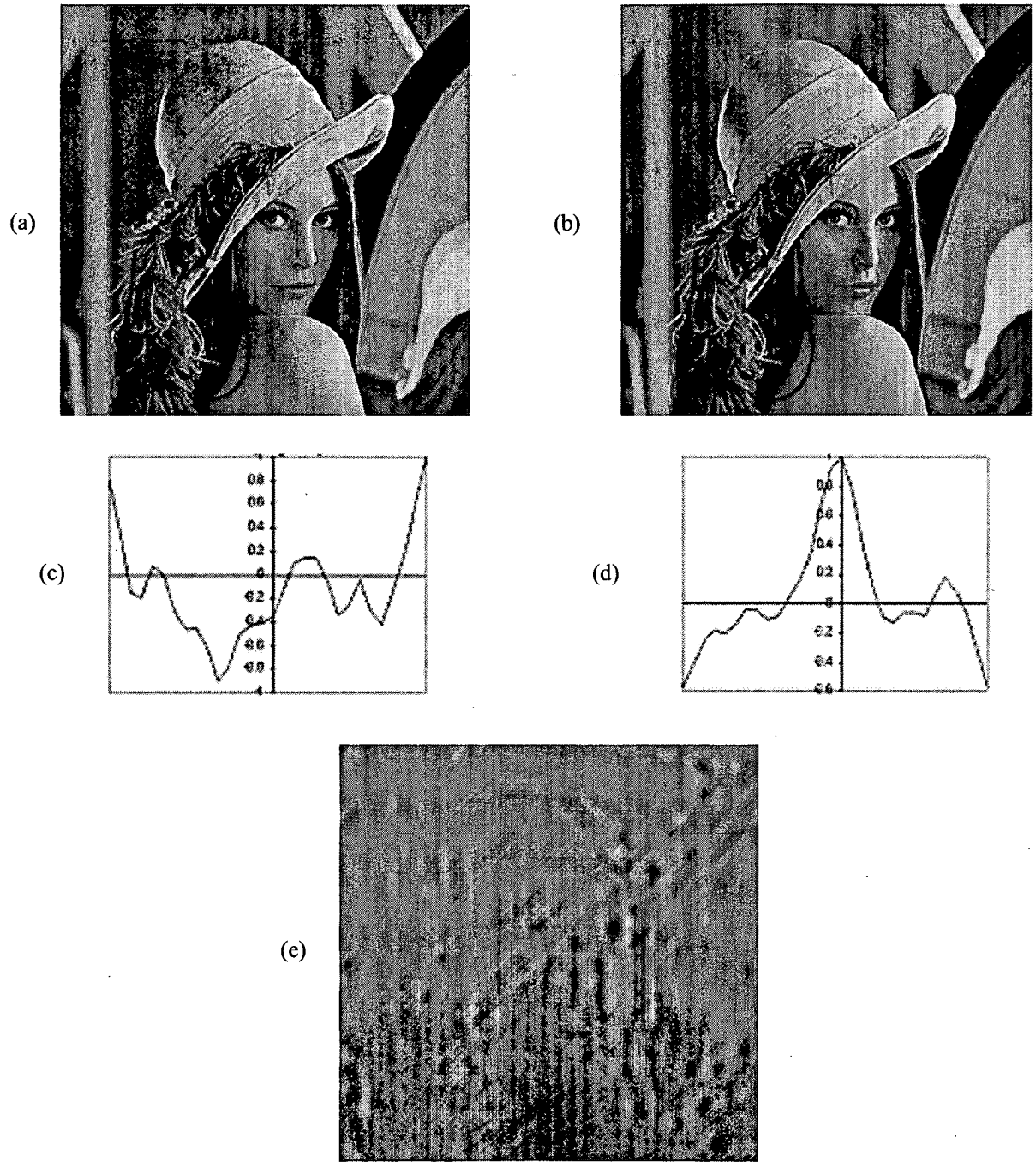


Figure 4: Results for Lena: (a) Original, (b) Watermarked one, (c) Correlation grafic for original, (d) Correlation grafic for watermarked, (e) Watermark.

Image Name	Optimal correlation	Random correlation 1	Random correlation 2	Random correlation3	Random correlation 4	Conclusion
Lena watermarked	**584609**	92605	133920	80534	143633	*watermarked*
Lena original	94538	98099	135492	76739	**137120**	*Non watermarked*

Figure 5: Results of correlation for Lena and decision.

making more information (e.g. ownership, date of marking) readable by the key owner from the watermark. This could be useful for real copyright protection protocols[8].[9]

8 REFERENCES

[1] Kahin B. The strategic environment for protecting multimedia. volume 1, pages 1–8. IMA Intellectual Property Project Proceedings, January 1994.

[2] Comes S. *Les traitements perceptifs d'images numérisées.* PhD thesis, Université Catholique de Louvain, June 1995.

[3] Olzak L.A. and Thomas J.P. Handbook of perception and human performance vol.1: Seeing spatial patterns. chapter 7.

[4] G.C. Phillips H.R. Wilson, D.K. McFarlane. Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23(9):873–847, 1983.

[5] G.C. Phillips H.R. Wilson. Orientation bandwiths of spatial mechanisms measured by masking. *J. Opt. Soc. Am. A*, 1(2):226–232, February 1984.

[6] Edited by Gustavus J. Simmons. Section 1: Chapter 4: 'public key cryptography' and section 2: Chapter 6: 'authentication: Digital signature' from 'contemporary cryptology: the science of information integrity' ieee press, 1992.

[7] J.F. Delaigle and C. De Vleeschouwer. Etiquetage d'images numériques en vue de la protection des droits d'auteur, Juin 1995.

[8] J.F. Delaigle C. Simon and B. Macq. Talisman (ac019): Technical state of the art. January 1996.

[9] O. Bruyndonckx J.M. Boucqueau and B. Macq. Watermarking: workpackage 5 of accopi. June 1995.